

Technologie sémantického webu pro vzdělávání

**Seminář VIRTUNIV, 5. května 2006
FI MU Brno**

Tomáš Gregar, Tomáš Pitner
{xgregar, tomp}@fi.muni.cz



- **Sémantický web**
(**Semantic Web, web se sémantikou**)
 - proč se o něm hovoří, co znamená (motivace, principy)
 - na čem se buduje (standards, technologie)
 - co přinese pro obecné použití a pro výuku

- **Na počátku počítače počítaly**
(angl. *computer*, něm. *Rechner*, čes. *počítač*)
 - pracují s čísly
- **Následně se velmi rychle uchytily jako stroje na zpracování informací/dat**
(franc. *ordinateur*)
 - pracují se strukturovanými daty
- **Čím více se dnes používají jako prostředek komunikace a sdílení (zejm. díky internetu)**
 - nutíme počítače pracovat i s daty semistrukturovanými/nestrukturovanými

Co je sémantický web (2)

- **Infrastruktura, kde počítače komunikují přes internet a „rozumějí“ přítomným informacím tam dostupným, se označuje jako *sémantický web*.**
- **Dnešní web je realizací distribuovaného informačního systému, kde SW komponenty jsou schopny vyměňovat si data a rozumět jejich *syntaxi*.**
- **SeWe je pokračováním webu stávajícího s důrazem na tzv. *sémantickou interoperabilitu*.**

Co je sémantický web (3)

- **Dnešní webový software (prohlížeče, autorské nástroje) produkuje a zpřístupňuje obsah srozumitelný na sémantické úrovni člověku, nikoli stroji (=dalšímu softwaru).**
- **Vše ostatní (tedy to podstatné - práce s vytvořenými a zpřístupněnými daty) je na člověku.**
- **Jen člověk dokáže v současnosti na základě dat na webu analyzovat problém, hledat řešení, rozhodovat...**

- **Stroje nad daty „neuvažují“**
- **Jediným prostředkem získání znalostí je nalezení vyhledávacími službami**
 - založené na *klíčových slovech* nebo
 - na *taxonomiích* („adresářích“) definovaných lidmi
 - je třeba se *umět ptát* (...a uživatelé to neumějí)
- **Vyhledávací služby**
 - někdy vrátí mnoho dokumentů, ale většinu irelevantních
 - jindy je dotaz příliš úzký či spíše voleny nevhodné termíny
 - závislé na jazyce (slovníku)
 - výsledky nejsou vidět v kontextu

- **Vyhledání je pouze první fází...**
 - udá lokaci potenciálně zajímavého zdroje, ale
 - nevyextrahuje z něj skutečně hledanou informaci a
 - výsledky často nejsou dále jednoduše strojově použitelné.

- **Tim Berners-Lee, *Business Model for the Semantic Web:***
 - "Now, miraculously, we have the Web. For the documents in our lives, everything is simple and smooth. But for data, we are still pre-Web."

- **Co by počítače v éře SeWe měly umět?**
 - Pochopit, co uživatel chce (napsat zprávu kolegovi, jehož e-mail neznám; naplánovat schůzku s dalšími lidmi, jejichž diáře nenosím v hlavě; najít mi cestu na ministerstvo do Prahy...)
 - Samy hledat relevantní informace (v záplavě informačních služeb najít tu relevantní, vědět, co požaduje za vstupy, tyto poskytnout a...)
 - Analyzovat informace (...následně rozumět výstupům služby, umět je zkombinovat a...)
 - Najít řešení (nabídnout formulář, kam stačí napsat zprávu a ta bude kolegovi odeslána; do mého mobilního diáře zapsat čas a místo schůzky; zobrazit mi v PDA cestu do Prahy...)

- **Základní vize SeWe - *Tim Berners-Lee* v časopisu *Scientific American*, r. 2001**

- **Formálněji (podle W3C)**

- Sémantický web je rozšířením současného webu, v němž informace mají přidělen dobře definovaný význam lépe umožňující počítačům a lidem spolupracovat.

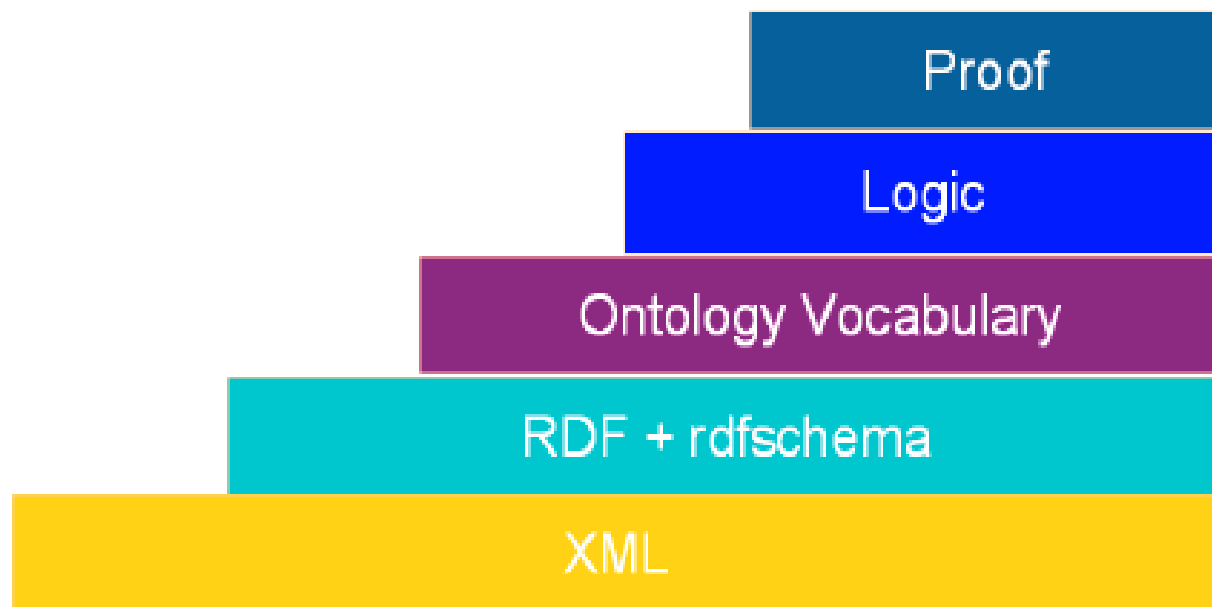
Sémantický web představuje reprezentaci dat na WWW.

Je založen na technologii *Resource Description*

Framework (RDF), která integruje širokou škálu aplikací využívajících *syntaktický zápis v XML a identifikátory URI pro pojmenovávání*.

- Jde tedy o to, aby data prezentovaná na internetu měla přesně definovaný význam a dovozovala do značné míry automatizované (strojové) zpracování.
- viz *Semantic Web*, <http://www.w3.org/2001/sw>

- **Vrstvy sémantického webu**
 - od dat přes informace ke znalostem



- **Správa znalostí (knowledge management)**

- podstatné nejen pro velké a distribuované firmy
- sdílení znalostí v rámci firmy
- správa vztahů se zákazníky, partnery
- objevování nových znalostí (např. pro rozhodování)
- nalézání případných nekonzistencí
- vizualizace informací
- správa práv k datům



- **B-to-C elektronický obchod**

- Business-to-Customer (firma -> zákazník)
- „přímý“ obchod s koncovým zákazníkem
- zveřejňování přesných a úplných informací o produktech
- automatizované vyhledávání optimálních nabídek
- možnost uplatnění automatických agentů při zprostředkování obchodních transakcí (nalezení nabídky, smlouvání, vlastní realizace transakce)

- **„Osobní agenti“**

- „Personal agents“
- automatické obstarání běžných denních úkonů osobního charakteru pomocí SW agentů
- dojednávání schůzek, management kontaktů,...

- **E-learning není pouhým elektronickým transferem výukového obsahu**
 - *"eLearning is just-in-time education integrated with high velocity value chains. It is the delivery of individualized, comprehensive, dynamic learning content in real time, aiding the development of communities of knowledge, linking learners and practitioners with experts" (Drucker 2000).*

- **Proč E-learning nad SeWe?**

- protože SeWe splňuje výše uvedené cíle e-learningu v oblastech
 - **transfer výukového obsahu** - SeWe umožňuje, aby i distribuovaně vytvářený obsah byl mapován na jednotící ontologie a tak systematicky zpřístupněn, lze tedy sestavovat „kurzy“ podle individuálních potřeb
 - **inteligentní agenti k řešení problémů** - podle momentálního kontextu učení dokáží agenti nad SeWe poskytnout rady, jak dál
 - **nelineární přístup k materiálům** - technologie SeWe umožňují vyhledat, zmapovat a vizuálně prezentovat i konceptuálně složité oblasti studovaného tématu - i bez didakticky předdefinovaného pořadí pojmů

- **Co dále SeWe nabízí**

- **symetrie** (ortogonalita): SeWe není orientován jen na e-learning, naopak, je integrující platformou vzdělávání s dalšími procesy v organizaci
- **distribuovaná tvorba obsahu**: *učitel* (expert, autorita) *není jediným zdrojem obsahu*, naopak, na vytváření se masivně podílejí *studující* a obecně *komunita*
- **adaptivita a personalizace**: obsah je přizpůsobován potřebám a možnostem studujícího na základě „sémantických“ pravidel

- **To hlavní, co přináší SeWe e-learningu**
 - možnost přesně zachytit *obsah* výukového materiálu (přes doménové ontologie)
 - možnost přesně zasadit výukový materiál do *kontextu*
 - strukturovat jej do *logických celků* a mapovat je na pojmy
- **Kromě těchto zjevných oblastí nelze opomenout roli SeWe v podpoře kolaborativní stránky vzdělávání**
 - vytváření a přístup ke komunitám
 - kolaborativní tvorba obsahu
 - ošetření práv přístupu, autorských práv...

- **Konceptualizace dat**

- datům dostupným na internetu je třeba formálně dát sémantiku - přiřadit je k příslušným *konceptům*. Prostředkem konceptualizace jsou *ontologie*.

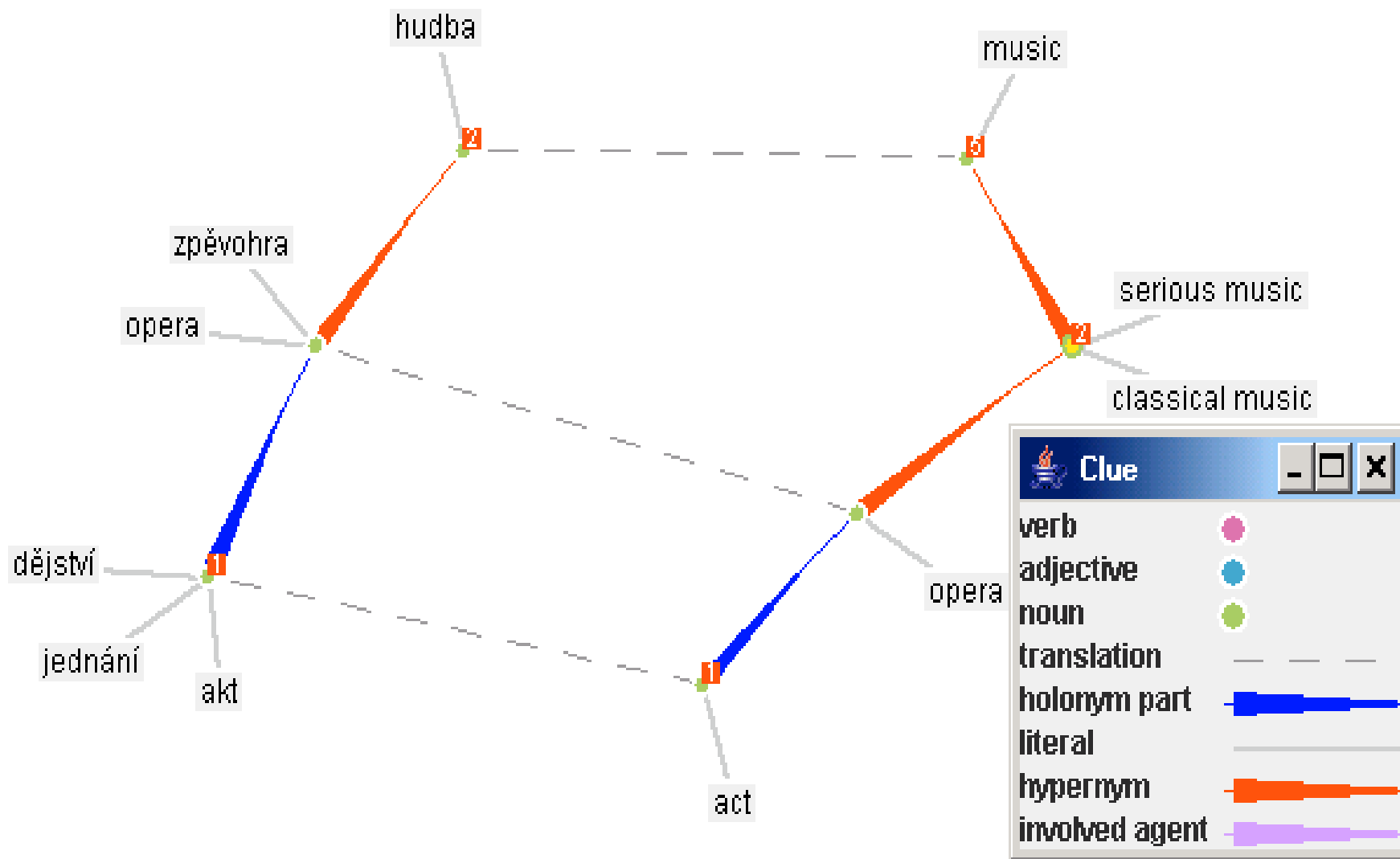
- **Ontologie**

- formalizované reprezentace znalostí určené k jejich sdílení a znovupoužití;
- jsou často doménového (oborového) zaměření a bývají konstruovány jako pojmové (konceptuální) hierarchie nebo sítě.

- **Ontologie terminologicky**
 - je tradičním filozofickým pojmem;
 - v informatice více používán v posledních deseti - patnácti letech (expertní systémy, web);
 - chápání relativně je ustáleno až nyní.
- **Ontologie prakticky - použití**
 - podpora porozumění v určité komunitě lidí (stejně chápání pojmů, umožnit odvozovat pojmy nové,...)
 - podpora komunikace a sdílení mezi počítači
- **Pro SeWe mají význam obě tyto oblasti**

- **Terminologické (lexikální) ontologie**
 - většinou hierarchické systémy pojmů a jejich reprezentací slovy/souslovími - termíny
 - obvyklými relacemi jsou nadtřída, podtřída (obecnější pojem, speciálněější pojem)
 - navíc často relace holonymie, meronymie (celek, část)
 - synonyma, antonyma... (shodný, protikladný význam)
 - rozpracováním myšlenky tezaurů
 - příklady: WordNet, EuroWordNet

Lexikální ontologie



- **Znalostní ontologie**

- kladou důraz na další (složitější) *vztahy* (relace) mezi *koncepty*

- **Pro SeWe mají význam oba typy**

- terminologické dávají základnu pro strojové rozpoznávání významu obsahu webu („porozumění“)
- znalostní dávají možnost nad významem pracovat (analyzovat, odvozovat další vztahy)

- **Doménové ontologie**
 - mapují pojmový aparát dané oblasti
 - dnes nejpoužívanější
- **Generické ontologie**
 - zachycují obecně platná fakta, vztahy
 - slouží mj. jako reference k připojení doménových ontologií (typicky v projektu *SUMO*)
- **Úlohové ontologie**
 - zaměřeny spíše na řešení konkrétních problémů zpracování znalostí než na obecná fakta
 - důraz na možnost odvozování a vývoj ontologie

- **Koncepty**
 - třídy, kategorie, rámce
- **Individua**
 - instance tříd
- **Relace**
 - vztahy, funkce, sloty, vlastnosti, atributy
- **Primitivní hodnoty**
 - řetězce, čísla... hodnoty relací (slotů)
- **Axiomy**
 - pravidla

- **Jak reprezentovat ontologii?**
 - vybrat správnou nosnou „logiku“
 - dostatečně silná, s dostupnými nástroji, pochopitelná pro cílovou skupinu uživatelů
 - ontologii zachytit pomocí specializovaných formálních i semiformálních jazyků odpovídajících zvolené logice
- **Jak zajistit vazbu konceptu určité ontologie na údaj, který je instancí tohoto konceptu?**
 - pomocí *metadat*

- **Metadata**

- data o datech;
- metaúdaj o určitém zdroji (např. webové stránce, položce v kalendáři...) může nést kýženou vazbu na konceptualizaci - ontologii v dané oblasti;
- metadata ovšem mohou sloužit i k jiným, techničtějším, účelům, např. zajistit interoperabilitu (např. údaje o formátu dat, kódování, výměnném protokolu).

- **Metadata byla využívána dávno před sémantickým webem...**

- ten z nich udělal svůj základní prostředek

- **Mezinárodní standardy**
 - ISO - std. výboru ISO/TC211
 - Dublin Core (**Dublin Core Metadata Initiative**)
<http://dublincore.org>
- **Národní MD std. veř. správy**
 - AGLS (Austrálie), GILS (USA), FGDC (USA)
- **Příklady ostatních MD std.**
 - IMS, SCORM (e-learning)
 - GDL (dig. knihovny)
 - ebXML (e-business)

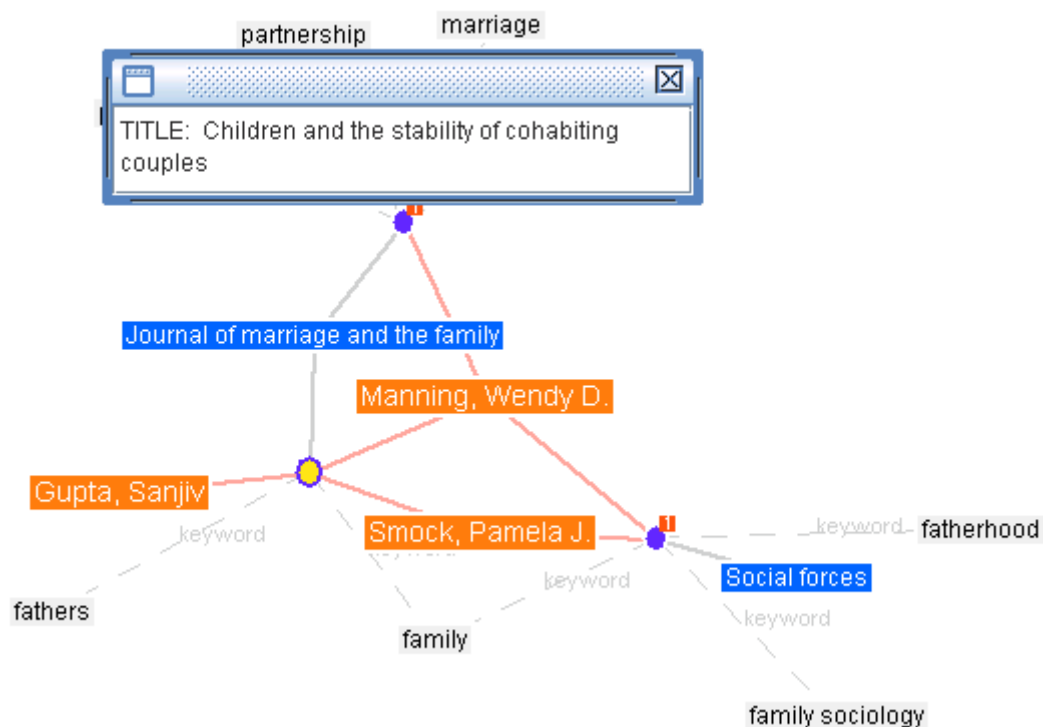
- **Dublin Core**

- jednoduchý standard metadat k (nejen) digitálním zdrojům;
- vyvinut knihovníky, dnes používán i pro webové zdroje;
- řada států používá DC + aplikační profil pro e-Government
- 15 základních prvků (DC Elements)


• **Metadatové prvky Dublin Core**

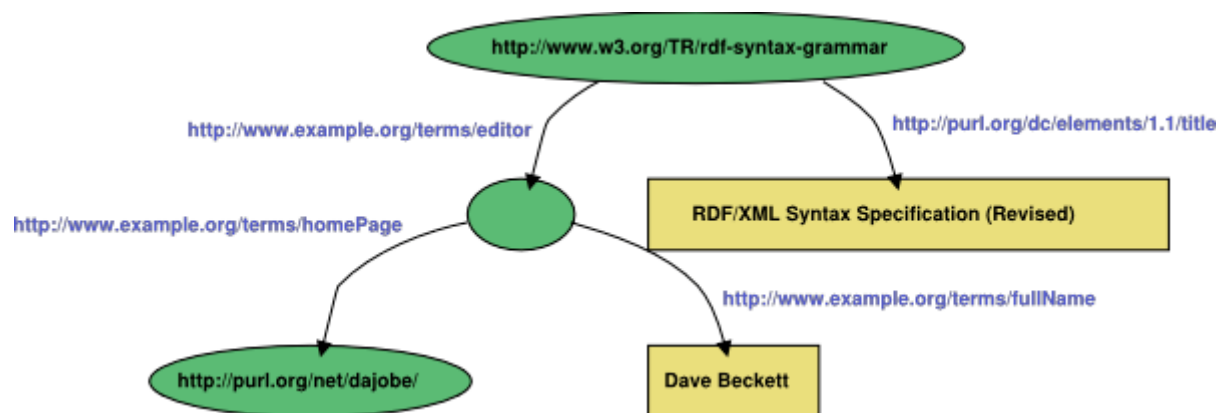
- **Název**
- **Tvůrce**
- **Předmět a klíčová slova**
- **Popis**
- **Vydavatel**
- **Příspěvatel**
- **Datum**
- **Typ zdroje**
- **Formát**
- **Identifikátor zdroje**
- **Zdroj**
- **Jazyk**
- **Vztah**
- **Pokrytí**
- **Správa autorských práv**

- Název, autor a klíčová slova publikací**



- **RDF (Resource Description Framework)**

- doporučení (standard) W3C 
- jednoduchý rámec pro modelování velmi obecných metadat



- **RDFS (Vocabulary Description Language 1.0: RDF Schema)**
 - doporučení (standard) W3C
 - jazyk na specifikaci modelů (schémat) RDF metadat pro konkrétní situace/oblasti
 - vymezuje přípustné vlastnosti a hodnoty pro určité třídy zdrojů/objektů

- **RDFS - příklady vlastností**

- rdfs:subClassOf
- rdfs:domain
- rdfs:label
- rdfs:isDefinedBy
- rdfs:member
- ...

- **OWL**

```
<owl:AllDifferent>
```

```
  <owl:distinctMembers rdf:parseType="Collection">
```

```
    <Opera rdf:about="#Don_Giovanni"/>
```

```
    <Opera rdf:about="#Nozze_di_Figaro"/>
```

```
    <Opera rdf:about="#Cosi_fan_tutte"/>
```

```
    <Opera rdf:about="#Tosca"/>
```

```
    <Opera rdf:about="#Turandot"/>
```

```
    <Opera rdf:about="#Salome"/>
```

```
  </owl:distinctMembers>
```

```
</owl:AllDifferent>
```

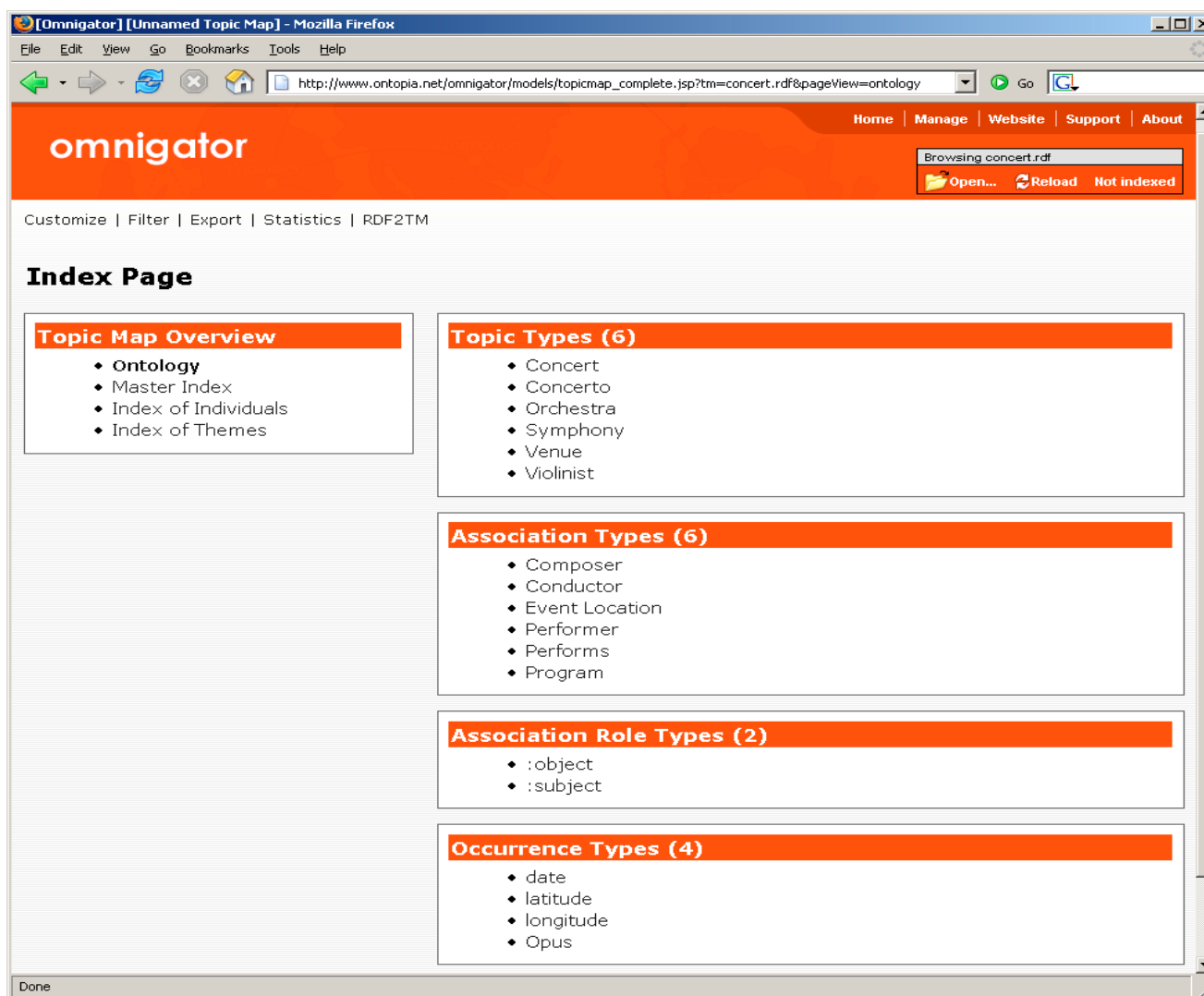
- **Metadata zprostředkovávají vazbu dat na „infrastrukturu“ sémantického webu**
 - např. v HTML stránce je metadata vyznačena vazba na příslušné pojmy (uzly) určité ontologie;
 - samotná ontologie může být zachycena podobnými prostředky jako metadata sama.

- **Popisy výukových zdrojů**

- data o výukových zdrojích, metadata;
- dnes již vysoce standardizované;
- to, že se nepoužívají na 100 % je dáno
 - dynamikou vývoje
 - používáním různorodých nástrojů
 - izolovaností - omezením na svět e-learningu
 - jejich postradatelností
 - nedůsledností při vytváření obsahu
 - nesnadnost znovupoužití výukových objektů snižuje motivaci k většímu využití standardních formátů, popisů

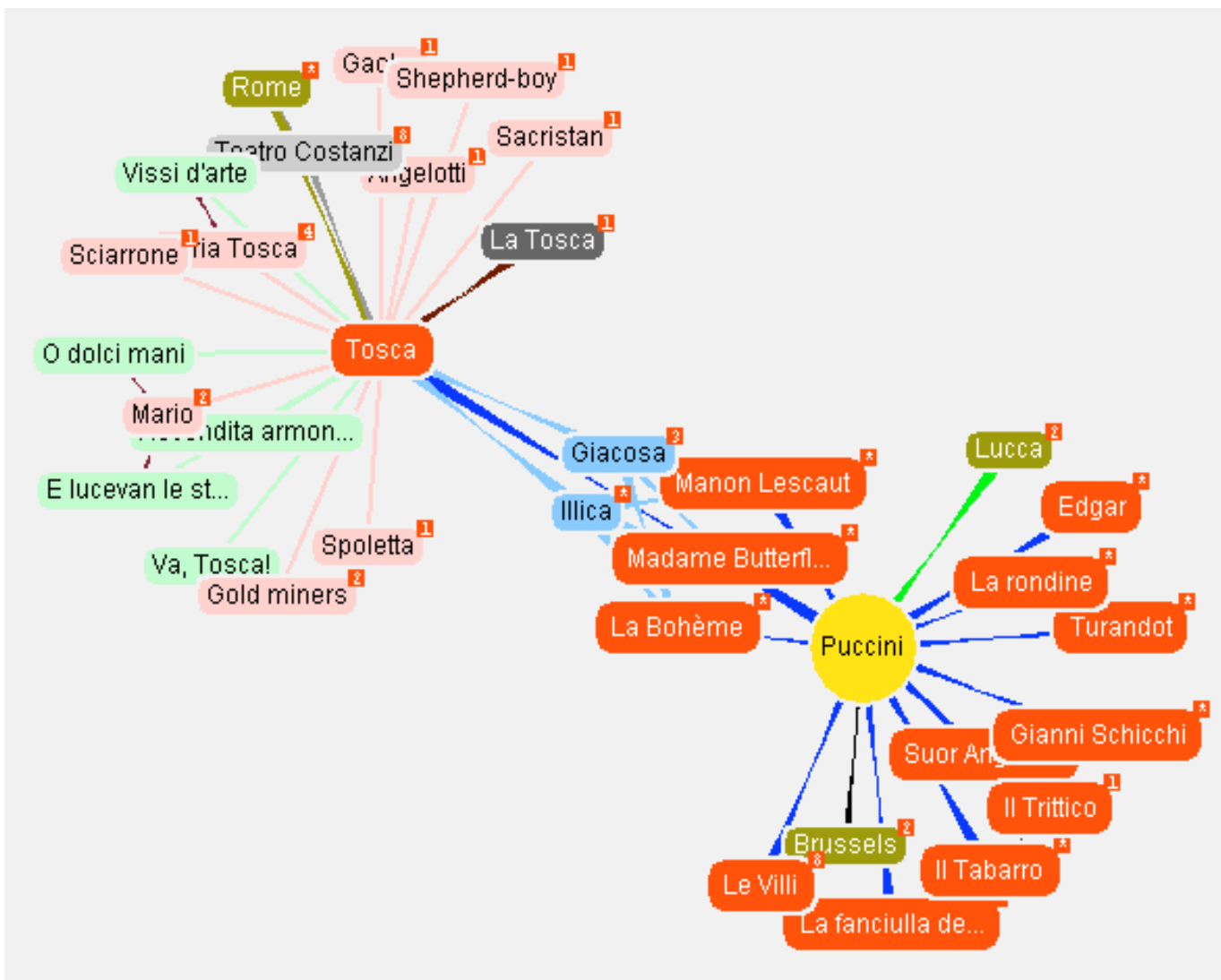
- **SCORM (vytvořen ADL)**
 - zastřešující standard metadatových popisů a způsobů balení a distribuce výukového obsahu
 - hlavní problém: „mělkost“ standardu - není zaručena skutečná znovupoužitelnost sdílených objektů
- **IEEE LOM**
- **AICC CMI**

- **Mapy témat (Topic Maps)**
 - jednodušší podoba zachycení ontologie
 - vhodné pro webová data/zdroje
 - standard W3C
 - méně obecnosti oproti modelům nad RDF
 - snazší strojové zpracování



The screenshot shows the Omnigator web application interface. The browser window title is "[Omnigator] [Unnamed Topic Map] - Mozilla Firefox". The address bar shows the URL: http://www.ontopia.net/omnigator/models/topicmap_complete.jsp?tm=concert.rdf&pageView=ontology. The page has a navigation menu with links: Home, Manage, Website, Support, About. The main content area is titled "Index Page" and contains several sections:

- Topic Map Overview**
 - **Ontology**
 - Master Index
 - Index of Individuals
 - Index of Themes
- Topic Types (6)**
 - Concert
 - Concerto
 - Orchestra
 - Symphony
 - Venue
 - Violinist
- Association Types (6)**
 - Composer
 - Conductor
 - Event Location
 - Performer
 - Performs
 - Program
- Association Role Types (2)**
 - :object
 - :subject
- Occurrence Types (4)**
 - date
 - latitude
 - longitude
 - Opus



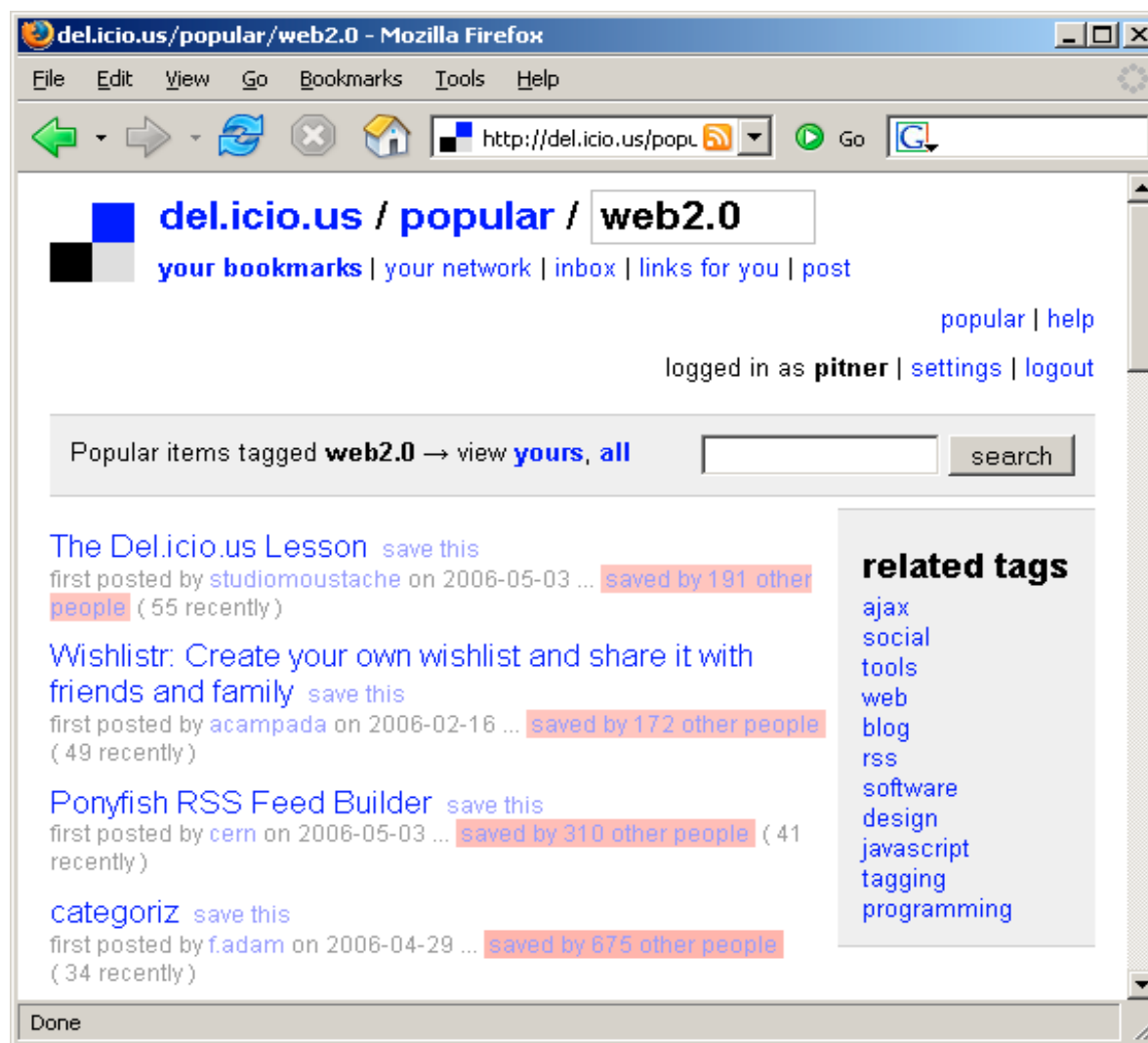
- **Co je Web 2.0?**

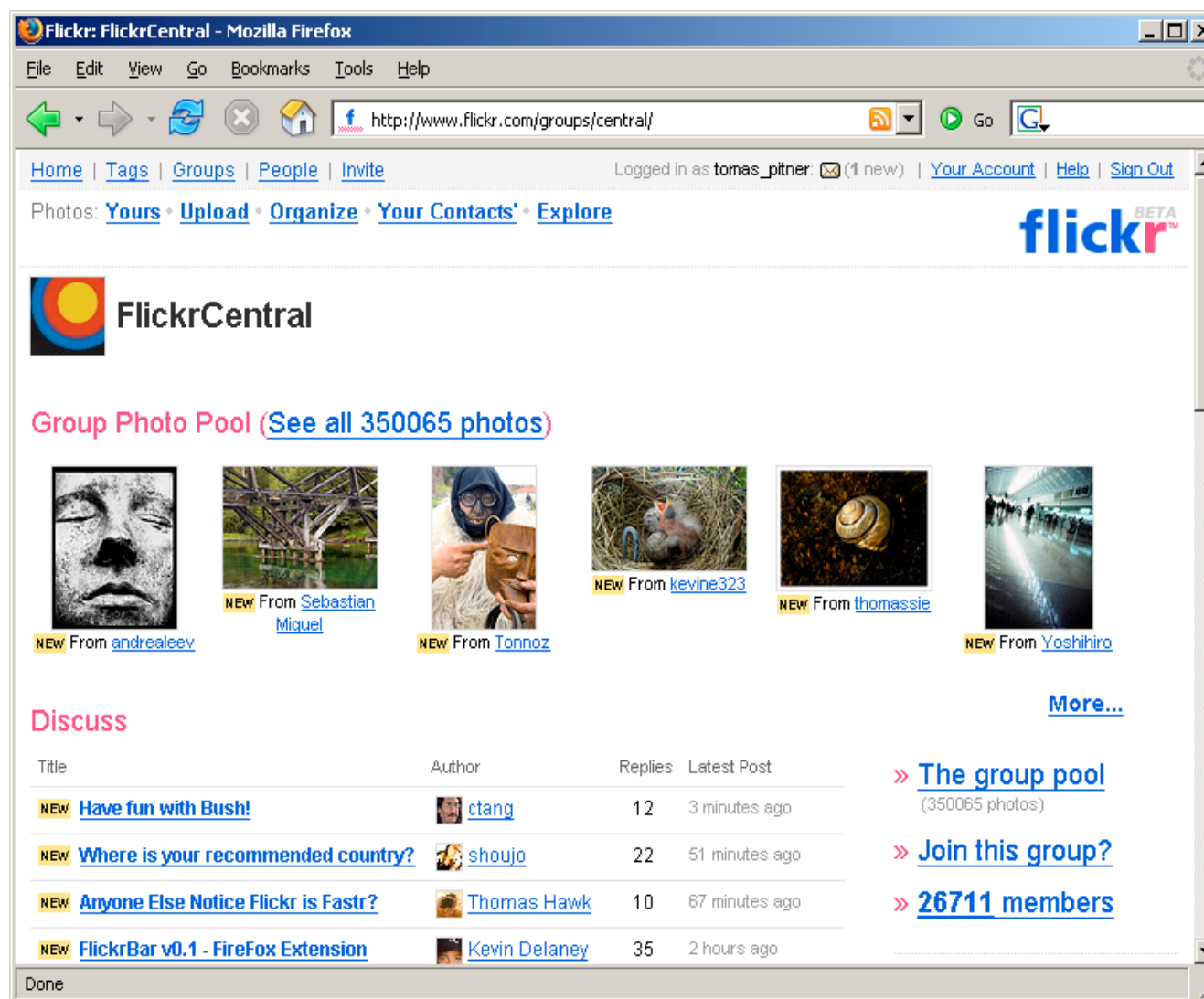
- trend v *technologiích*, ale hlavně v *užívání* webu;
- jednou větou těžko definovatelný pojem;
- lze charakterizovat jako přístup k chápání webových služeb jako komponent *globálního informačního systému*;
- důraz je kladen na *komunikaci a sdílení*;
- stírá se rozdíl mezi *poskytovatelem a příjemcem* informací, web se stává „read-write“;

- **Web 2.0 znamená**

- důraz na služby, ne hotové produkty -> vše se může rychleji vyvíjet;
- služby musejí být efektivně škálovatelné;
- chápou web jako platformu, ne jako distribuční kanál;
- využívají obvykle síťový efekt (více uživatelů -> větší kvalita);
- uživatelé se a priori důvěřují (wiki, weblogging);
- očekávané využití služby se nemusí shodovat s původním záměrem;
- jemná adresovatelnost zdrojů (tím může být i „drobnost“: záznam v blogu, položka v diáři, vizitka).;

- **Google (Search, Maps, Gmail, Calendar, AdSense)**
 - služby využívají bohaté, globální, „neomezené“ datové báze
 - silně interagují s uživatelem
 - používají (aspoň částečně) technologie SeWe (zpracování přirozeného jazyka)
 - jsou relativně dobře integrovatelné, existují k nim API
 - kladou důraz na práci v komunitě (samotný Gmail je komunita)
- **Mapy.cz**
 - vysoce interaktivní, chová se jako desktopová aplikace
- **Del.icio.us**
 - systém na evidenci a sdílení zajímavých odkazů
- **Flickr.com**
 - sdílitelná alba fotografií s mnoha funkcemi





Flickr: FlickrCentral - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.flickr.com/groups/central/

Home | Tags | Groups | People | Invite Logged in as tomas_pitner (1 new) | Your Account | Help | Sign Out

Photos: [Yours](#) • [Upload](#) • [Organize](#) • [Your Contacts](#) • [Explore](#)

flickr BETA

FlickrCentral

Group Photo Pool ([See all 350065 photos](#))

NEW From [andrealeev](#) NEW From [Sebastian Miquel](#) NEW From [Tonnoz](#) NEW From [kevine323](#) NEW From [thomassie](#) NEW From [Yoshihiro](#)

Discuss [More...](#)

Title	Author	Replies	Latest Post
NEW Have fun with Bush!	ctang	12	3 minutes ago
NEW Where is your recommended country?	shoujo	22	51 minutes ago
NEW Anyone Else Notice Flickr is Fastr?	Thomas Hawk	10	67 minutes ago
NEW FlickrBar v0.1 - FireFox Extension	Kevin Delaney	35	2 hours ago

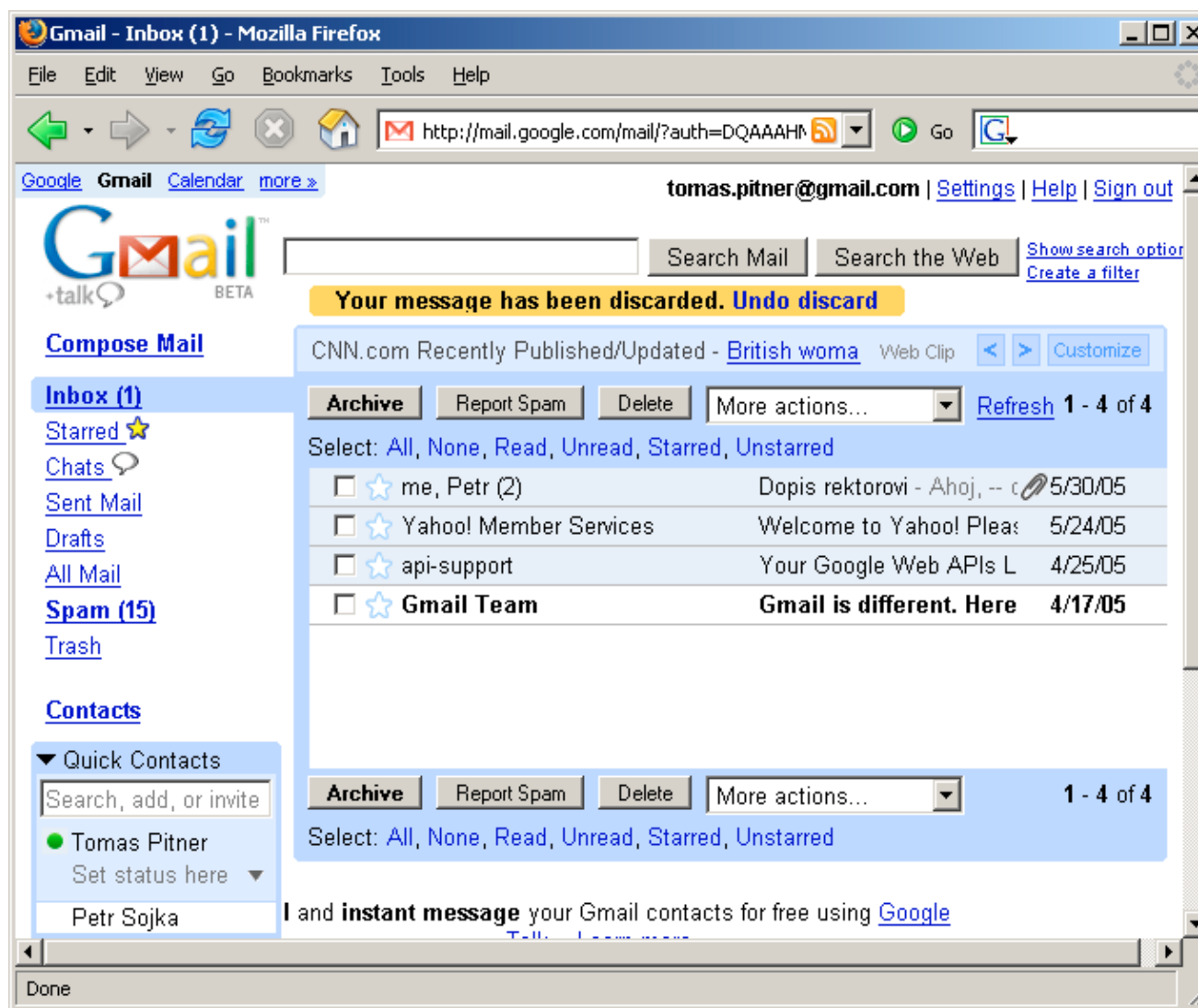
» [The group pool](#) (350065 photos)

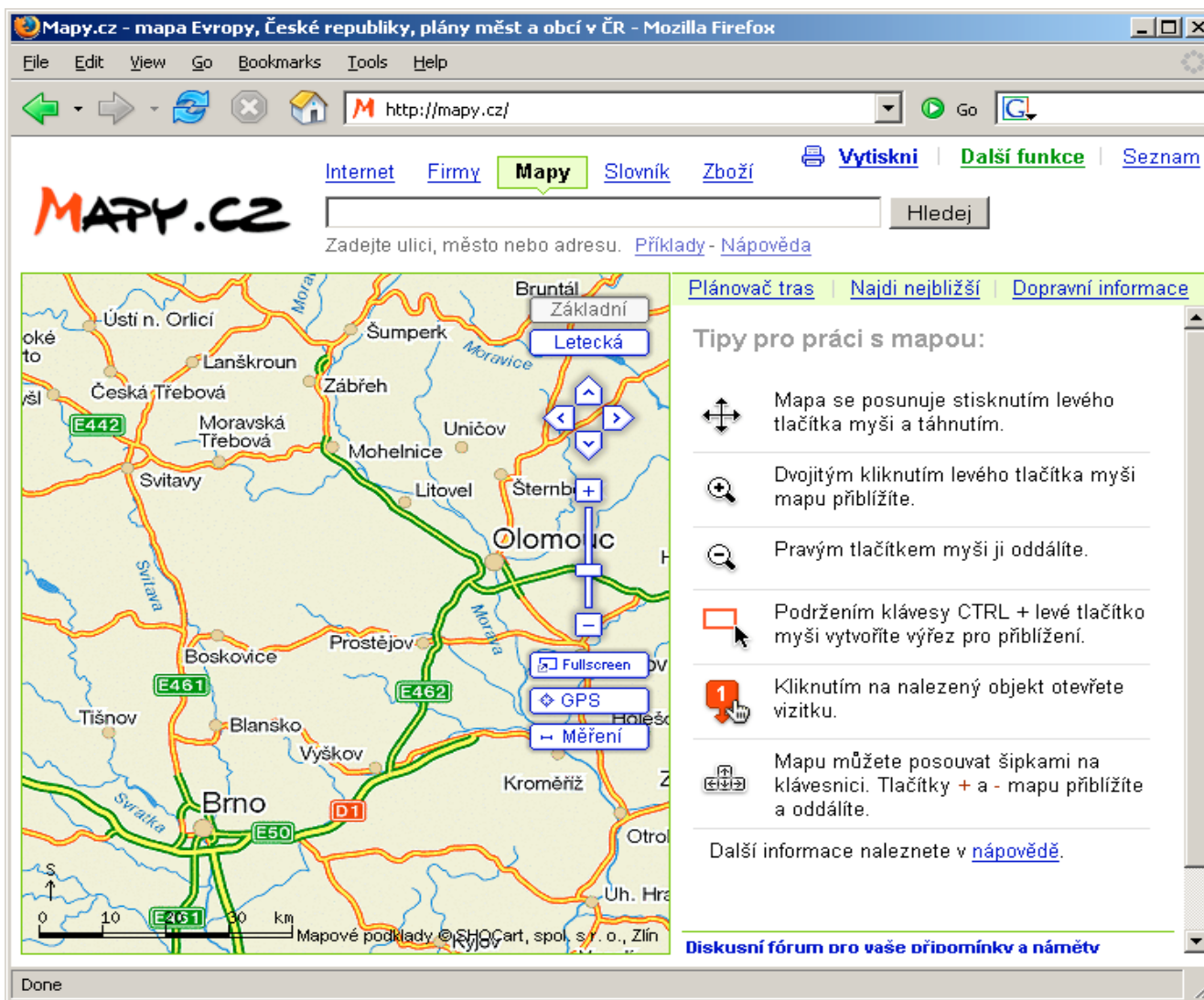
» [Join this group?](#)

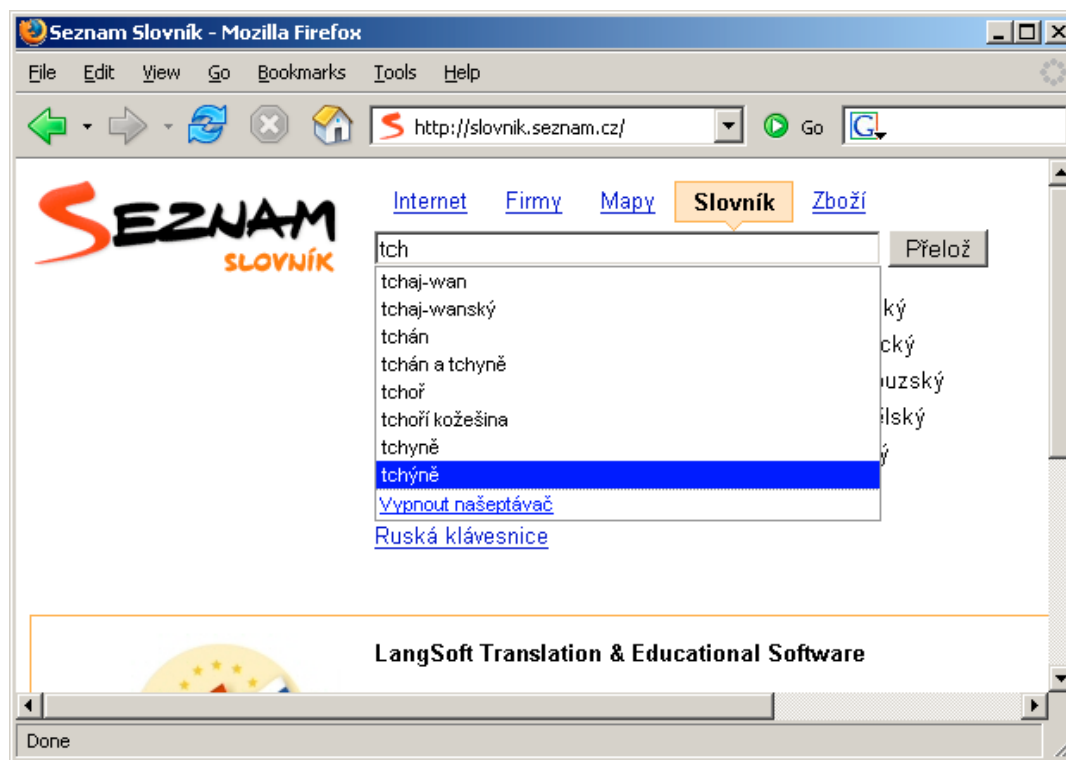
» [26711 members](#)

Done

- **Protokol**
 - HTTP
- **Výměnný formát**
 - XML
- **Obslužný SW**
 - na serveru „cokoli“ nad HTTP/XML
 - na klientovi AJAX (Asynchronous JavaScript with XML)
- **Vlastně nic technologicky převratného - vše existovalo, jen se důsledně nevyužívalo, spíše obcházel...**







- **Web 2.0 lze tedy realizovat se stávající technologickou výbavou, dodržují-li se některé zásady**
- **Příklad architektonických doporučení typických pro Web 2.0 - REST (REpresentational State Transfer)**
 - pojem zaveden v disertaci R. Fieldinga, r. 2000;
 - jakýsi „návrat ke kořenům webu“;
 - popisuje architektonické zásady umožňující využít skutečný potenciál webu - jednoduchost, interoperabilita;

- **Stavební kameny REST architektury**

- každá služba (zdroj, *resource*) má svou identifikaci (URI)
- zdroje mají své reprezentace (*representation*) - HTML, XML, obrazová data...;
- služby jsou *bezstavové*, stav se *vyměňuje* při komunikaci v zasílaných reprezentacích (odtud *Representational State Transfer*) ;
- budováno nad základním HTTP se všemi výhodami i negativy (možnost kešování, služba může být současně klientem i serverem, lze snadno integrovat...).

- **Web 2.0 zahrnuje do svého potenciálu i „sémantické technologie“**
 - chápe je méně rigidně, než se původně uvažovalo;
 - SeWe se prosadí, ale postupně, nenásilně, v podobě Webu 2.0 (nebo až 3.0 ???);
 - každý uživatel Webu 2.0 je přispěvatelem, těžko ho k něčemu nutit;
 - formální postupy SeWe (např. budování univerzálních taxonomií) se zatím neuplatňují;
- **To neznamená SeWe nyní nechat odpočívat!**
 - principy a možnosti je třeba poznat a ty relevantní používat!